

# 基于注意模型的视觉替代方法

田亚男,雷红玮,王 旭

(东北大学信息科学与工程学院,辽宁沈阳 110819)

**摘 要:** 现有的视觉替代方法均是在特定环境中,通过目标识别获得映射特征,缺少普遍的适用性.针对这一问题,提出了一种基于注意模型的视觉替代方法.利用人眼的视觉特性,提取图像的感兴趣区域,并根据听觉显示的基本原理,提出了将感兴趣区域的位置、尺寸和颜色映射为音符响度、音长和音调的PSC映射方法.实验结果表明:将图像中引起视觉注意的区域映射为电子音符,符合人类视觉认知过程,有利于盲人获得外部环境的重要信息,降低了盲人训练和学习的难度,并且听感良好,不会造成听觉疲劳.

**关键词:** 视觉替代; 听觉显示; 视觉注意; 感兴趣区域

**中图分类号:** TP391.41      **文献标识码:** A      **文章编号:** 0372-2112 (2014)05-0890-06

**电子学报 URL:** <http://www.ejournal.org.cn>

**DOI:** 10.3969/j.issn.0372-2112.2014.05.009

## A Vision Substitution Method Based on Visual Attention Models

TIAN Ya-nan, LEI Hong-wei, WANG Xu

(School of Information Science and Engineering, Northeastern University, Shenyang, Liaoning 110819, China)

**Abstract:** The existing vision substitution methods are only used in special environments because their mapping features are obtained by object recognition. Aiming at this problem, a vision substitution method based on visual attention models is proposed. Firstly, regions of interest are extracted according to human visual properties. And on the basis of the auditory display theory, PSC mapping method is presented that the position, size and color of regions of interest are respectively mapped into note loudness, duration and pitch. Results show that this method corresponds with the cognitive processes of human vision. It is good at helping the blind to obtain the important information and reducing the difficulty of training and learning. Moreover, the mapped electronic music is pleasing to ear and enjoyable.

**Key words:** vision substitution; auditory display; visual attention; region of interest (ROI)

### 1 引言

视觉是人类获得外部环境信息的主要途径,视觉的损伤给人类的生活带来极大的不幸.据2003年世界卫生组织公布的数据,全世界目前约有4500万人目盲,1.35亿人视力低下.随着人口的老齡化,盲人和视力低下的人数还会不断攀升.因此研究人员一直在探索用其他的感觉器官来获得视觉信息,即视觉替代问题.

根据视觉替代中图像信息的映射形式,可以将视觉替代方法分为基于像素的方法和基于目标识别的方法.基于像素的方法将环境图像中的每个像素的位置和灰度等特征直接映射为不同的声音信号.1992年Meijer等人提出的Voice系统采用的就是基于像素的方法.该系统对灰度图像像素从左到右逐列扫描,将图像顶部映射成高频音调,图像底部映射成低频音调,像素值直接转

换成声音的强度<sup>[1]</sup>.这种方法映射简单、易于实现.但是由于该方法将图像中全部像素进行映射,包含了大量的背景信息,盲人需要经过长时间的学习训练,才能够理解图像的内容.基于目标识别的方法将图像中的目标物识别出来,将目标物的尺寸、位置、形状等特征映射为声音信号.该方法去除了背景的影响,降低了盲人学习训练的难度.例如Sainarayanan等人2004年提出的通过识别前方的障碍物来辅助盲人独立行走的NAVI系统<sup>[2]</sup>; Bologna等人2009年提出的通过识别路面上预先绘制的红色导行线来帮助盲人独立行走的Serpentine系统<sup>[3]</sup>; 以及Chen等人2011年提出的基于动态模糊彩色分割的视觉替代系统<sup>[4]</sup>.尽管基于目标识别的视觉替代方法取得了较好的实验结果,但是他们只能在特定的环境中使用,没有普遍的适用性.这是因为在不同的环境中,目标物的特征不同,无法建立统一的目标识别模型.并且

他们都没有考虑到人类视觉对环境图像中不同区域的选择机制.众所周知,图像需要由人的视觉系统来感知和解释.因此在视觉替代问题中,可以通过视觉感知特性来选择图像特征进行映射.

本文提出了一种基于注意模型的视觉替代方法,利用人眼的视觉特性,提取图像的感兴趣区域(Region Of Interest, ROI),并根据听觉显示的基本原理,提出了将感兴趣区域的位置、尺寸和颜色映射为音符响度、音长和音调的 PSC 映射方法.图像的感兴趣区域包含了图像中重要的内容信息,盲人通过听音获得这些信息,从而帮助他们行走和感知外界环境.

## 2 视觉注意模型

视觉注意机制的研究表明:人眼对信息的处理不是均衡的,通过眼球的扫视,将感兴趣的区域移动到具有高分辨率的视网膜中央凹区,从而实现对该区域的注意. Itti 等人提出的基于显著度的视觉注意模型,较好的模拟了人眼观察静态图像时注意的选取过程.使用线性高斯滤波器对输入图像进行低通滤波,再通过下采样得到一系列不同尺度的图像,构成高斯金字塔.低尺度图像中的每个像素均可以看作是与原图位置相对应的区域.中心点与其周围区域的反差可以通过中心-周边算子对处于金字塔不同层的图像差减来计算,即

$$f(c, s) = |f(c) - f(s)| \quad (1)$$

其中  $f$  表示高斯金字塔中不同尺度的图像;  $c$  表示中心尺度,  $s$  表示边缘区域的尺度.定义为两个不同尺度图像之间的差减操作.

算法的具体步骤如下:

### 步骤 1 生成特征图

首先,从输入图像中提取颜色、亮度和方向特征.设  $r, g, b$  分别为输入图像的红绿蓝分量.用  $r, g, b$  的平均值定义图像的亮度特征,即  $I = (r + g + b)/3$ .用亮度图像  $I$  生成亮度高斯金字塔  $I(\sigma)$ ,其中  $\sigma \in [0..8]$ .对  $I(\sigma)$  应用中心-周边算子,得到 6 幅亮度特征图,即  $I(c, s) = |I(c) - I(s)|$ ,  $c \in \{2, 3, 4\}$ ,  $s = c + \delta$ ,  $\delta \in \{3, 4\}$ .

为了在亮度特征中增强颜色的影响,用亮度  $I$  归一化  $r, g, b$  通道.由于颜色的改变在低亮度时很难被人眼察觉,同时为了避免  $r, g, b$  分量在低亮度时出现大幅波动,所以只有在该像素的亮度大于图像最大亮度值的 10% 时,执行归一化操作,否则直接将  $r, g, b$  分量置为零.应用式(2) - (5)将  $r, g, b$  通道进一步分解为四个颜色通道,且当  $R, G, B, Y$  为负值时置零.由这四个颜色通道生成四个高斯金字塔  $R(\sigma), G(\sigma), B(\sigma)$  和  $Y(\sigma)$ .

$$\text{红色: } R = r - (g + b)/2 \quad (2)$$

$$\text{绿色: } G = g - (r + b)/2 \quad (3)$$

$$\text{蓝色: } B = b - (r + g)/2 \quad (4)$$

$$\text{黄色: } Y = (r + g)/2 - |r - g|/2 - b \quad (5)$$

在人类大脑皮层接收区域的中心,神经元对一种颜色敏感(如红色)的同时会抑制其对比颜色(如绿色).这种空间和颜色的对比存在于红/绿、绿/红、蓝/黄、黄/蓝颜色对中.因此,颜色高斯金字塔的每一层进一步被分解为红-绿(RG)和蓝-黄(BY)颜色对比图.对颜色对比图应用中心-周边算子,如式(6)和(7)所示,得到 12 幅颜色特征图.

$$RG(c, s) = |(R(c) - G(c))(G(s) - R(s))| \quad (6)$$

$$BY(c, s) = |(B(c) - Y(c))(Y(s) - B(s))| \quad (7)$$

通过对亮度图  $I$  进行方向 Gabor 金字塔分解,获得局部方向信息  $O(\sigma, \theta)$ ,这里  $\sigma \in [0..8]$  表示金字塔的各图层,方向参数  $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ .对  $O(\sigma, \theta)$  应用中心-周边算子,如式(8)所示,获得 24 幅方向特征图.

$$O(c, s, \theta) = |O(c, \theta) - O(s, \theta)| \quad (8)$$

### 步骤 2 将特征图融合成一幅显著图

经过步骤 1 的处理,共获得了 42 幅特征图.这些特征图表示了视觉不同的早期反应,具有不同的变化幅度和构造机制.为了防止在少数特征图中有强烈表现的特征被其他大量不显著的特征所掩盖,在图像融合之前执行  $N(\cdot)$  归一化操作,将有少量强活动峰值的特征图整体提升,将有大量同样峰值响应的特征图整体下调.归一化后的特征图,按照式(9) - (11)融合成亮度显著图  $\bar{I}$ ,颜色显著图  $\bar{C}$  和方向显著图  $\bar{O}$ .

$$\bar{I} = \bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c+4} N(I(c, s)) \quad (9)$$

$$\bar{C} = \bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c+4} [N(RG(c, s)) + N(BY(c, s))] \quad (10)$$

$$\bar{O} = \sum_{\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}} N[\bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c+4} N(O(c, s, \theta))] \quad (11)$$

其中  $\bigoplus$  表示跨尺度加法,通过将每一幅图像降低到尺度 4 后逐点相加来实现.最后,将亮度、颜色和方向显著图融合得到一幅最终的显著图,如式(12)所示.得到显著图  $S$  后,将  $S$  中显著性较强的区域确定为感兴趣区域<sup>[5-7]</sup>.

$$S = (N(\bar{I}) + N(\bar{C}) + N(\bar{O}))/3 \quad (12)$$

## 3 听觉显示

听觉显示是指利用声音来表示信息的技术.听觉显示具有下面两个优点:第一,听觉显示的载体——声音具有多维特性,适合表示多维数据,而不会造成数据的过载.对人耳的感觉来说,声音可以用音调、响度、音长等维度来描述,声音的频率决定音调,振幅决定响度,振动时间决定音长.第二,听觉具有并行输入特性,

可以进行信息的多任务监测和处理. 因此听觉显示在视觉替代的研究中得到了广泛应用. Voice 和 NAVI 系统都是听觉显示的视觉替代系统.

听觉显示的核心是设计一种图像与声音之间的映射方法. 本文提出了一种 PSC(位置、尺寸和颜色)映射方法. 该方法的基本思想是: 将 ROI 区域的水平位置映射为响度; 垂直位置映射为音高; 尺寸映射为音长, 主颜色映射为音调; 最终合成电子音符来表示 ROI 区域的主要特征.

### 3.1 位置和尺寸特征的映射

ROI 区域的位置特征可以用质心坐标  $(x_c, y_c)$  来表示, 如公式(13)所示, 其中  $f(i, j)$  为 ROI 区域灰度图像的像素强度,  $P, Q$  为 ROI 区域内水平方向和垂直方向的像素个数. 由于 ROI 区域中包含了引起视觉注意的目标的大部分和少量背景, 因此通过计算 ROI 区域的灰度质心, 可以获得目标的位置.

$$\begin{cases} x_c = \frac{\sum_i \sum_j i * f(i, j)}{\sum_i \sum_j f(i, j)} \\ y_c = \frac{\sum_i \sum_j j * f(i, j)}{\sum_i \sum_j f(i, j)} \end{cases} \quad (13)$$

将质心的水平位置  $x_c$  映射为响度, 垂直位置  $y_c$  映射为音高. 用 ROI 区域内的像素数目表示 ROI 区域的尺寸, 并将其映射为音长.

### 3.2 颜色特征的映射

颜色是 ROI 区域的重要特征之一. RGB 彩色图像一共可以表示 16777216 种不同的颜色. 而人类听觉系统可以听到的声音的音调在 20 赫兹到 20 千赫兹之间. 如果颜色和声音之间直接进行映射, 则两个不同颜色映射后音调差约为 0.0012 赫兹, 远小于人耳的分辨率. 并且映射的音调越多, 盲人学习和使用的难度就越大. 因此, 本文首先识别出 ROI 区域的主颜色, 然后将其映射为“十二平均律”设定的音符的频率.

在提取图像的主颜色时, 先将图像变换到更加符合人类色彩感知的 HSV(色调、饱和度、亮度)空间. 由于 HSV 空间的颜色信息非常丰富, 因此需要对 HSV 空间进行量化<sup>[8]</sup>. Smith 提出了一种在 HSV 空间上进行 166 色量化的方法<sup>[9]</sup>, 该方法基本符合人类视觉的感知特性, 在相同的量化区间内的像素可以近似认为是相同的颜色. 假设 HSV 空间的像素点  $c(i, j) = [h, s, v]$ , 用  $q(i, j) \in [0 \cdot 165]$  表示该像素点量化后的颜色值, 则:

$$q(i, j) = \begin{cases} 0, & v \leq 0.1 \\ g(h, s, v), & s < 0.1 \& v > 0.1 \\ y(h, s, v), & \text{其他} \end{cases} \quad (14)$$

其中

$$g(h, s, v) = \begin{cases} 1, & s < 0.1 \& 0.1 < v \leq 0.4 \\ 2, & s < 0.1 \& 0.4 < v \leq 0.7 \\ 3, & s < 0.1 \& 0.7 < v \leq 1 \end{cases} \quad (15)$$

$y(h, s, v)$  均匀划分剩余的 HSV 空间, 色调  $h$  以  $20^\circ$  间隔均分成 18 份, 饱和度  $s$  和亮度  $v$  以 0.3 为间隔均分成 3 份. 量化后整幅图像的颜色被划分为 166 个颜色区间, 相同区间的颜色可以用该区间的典型颜色值来表示.

研究表明, 人眼在识别颜色时, 主要将颜色分为红、绿、蓝、黄、紫、橙、粉、棕、灰、白、黑十一种颜色<sup>[10]</sup>. 因此, 根据人眼目视的结果, 将 166 色量化的典型颜色划分到上述十一种颜色中, 生成量化值与颜色的映射表. 如果直接对量化后的图像应用颜色直方图提取主颜色, 则会出现如下问题: 第一, 在光照不足、过分曝光等情况下, ROI 区域中的目标物上会形成一些颜色畸变的像素点, 或者目标物本身具有多种颜色, 当这些像素点和背景的颜色相同时, 可能会使颜色直方图法提取的主颜色与人眼判断的主颜色不同; 第二, ROI 区域是在不同尺度下通过融合亮度、颜色及方向显著图形成的, 因此当 ROI 区域内的目标物较小时, ROI 区域中的背景较多, 此时应用颜色直方图提取主颜色, 会将背景色作为主颜色, 这显然和人眼的视觉特性是不符的. 为了解决上述问题, 文中提出了一种基于颜色域的 FDH(滤波、膨胀及颜色直方图)算法, 用来提取 ROI 区域的主颜色. 算法如下: 设  $q(i, j)$  为图像的 166 色量化矩阵,  $R$  为  $q(i, j)$  中 ROI 区域的模板, 计算模板内各个颜色量化值出现的次数, 即:

$$N = \{ \langle n[c_1], n[c_2], \dots, n[c_k], \dots, n[c_m] \rangle \mid n[c_k] > 0 \} \quad (16)$$

其中  $n[c_k]$  表示颜色量化值  $c_k$  出现的次数. 将模板中心对应的量化值赋值为模板中出现次数最多的量化值, 获得颜色域滤波后的量化值矩阵  $\tilde{q}(i, j)$ , 即:

$$\tilde{q}(i, j) = \{ c_k \mid n[c_k] = \max\{N\} \} \quad (17)$$

对 ROI 区域中的所有量化值执行上述操作后, 能够去除图像中像素数目较少的颜色突变点, 实现颜色域的滤波, 从而提高主颜色提取的准确率.

通过对 ROI 区域内目标物分布情况的分析可知: 目标物通常位于 ROI 区域的中心部分, 并占据一定的面积; ROI 区域的质心  $(x_c, y_c)$  落在目标物上. 而在目标物的像素数目和背景像素数目接近或者略小于背景像素数目时, 人眼仍然将中心目标物的颜色判断为 ROI 区域的主颜色. 针对人眼的这种特性, 本文提出了一种基于颜色域的膨胀算法: 对于颜色量化矩阵  $\tilde{q}(i, j)$ , 颜色域的膨胀定义为:

$$D(s, t) = \{ \tilde{q}(i, j) \mid \tilde{q}(i, j) \cap b(s, t) \neq \emptyset \} \quad (18)$$

其中  $b(s, t)$  为结构元素. 通过不断移动结构元素, 将中心目标物向四周膨胀. 首先, 以 ROI 区域质心为原点, 将 ROI 区域划分为四个象限; 然后, 对不同的象限以及通过质心的行和列, 分别使用图 1 所示的结构元素进行颜色域的膨胀操作, 图中“+”表示

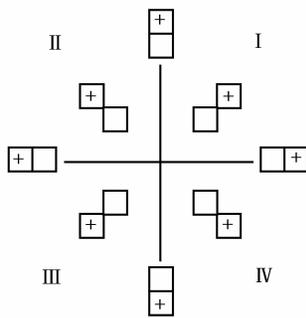


图1 颜色域膨胀使用的结构元素

结构元素的参考点. 多次执行膨胀操作后, ROI 区域的目标物向四周扩张, 背景被缩减. 最后, 对膨胀后的颜色量化值矩阵执行颜色直方图操作, 统计每个颜色量化值出现的次数, 按照量化值与颜色的映射表将颜色量化值划分到十一种颜色中, 将像素数最多的颜色作为 ROI 区域的主颜色. 按照表 1 所示的映射关系, 将 ROI 区域的主颜色映射为特定的音符.

表 1 颜色与音符 (C4-B4) 的映射关系

颜色	音符	颜色	音符	颜色	音符
白色	C4	黄色	E4	紫色	A4
红色	# C4	灰色	F4	粉色	# A4
棕色	D4	绿色	G4	黑色	B4
橙色	# D4	蓝色	# G4		

### 3.3 电子音符的形成

获得 ROI 区域的位置、尺寸和颜色特征后, 根据 PSC 映射方法的基本思想, 将 ROI 区域质心的水平方向坐标  $x_c$  映射为响度; 垂直方向坐标  $y_c$  映射为音高; 尺寸映射为音长, 主颜色映射为音调. 为了增强响度变化的可识别性, 采用按指数增加的三种响度等级. 选择相邻的三个八度音符 C4-B4、C5-B5 和 C6-B6 表示不同的音高. 将 ROI 区域内像素点的个数映射为音长, ROI 区域越大音符播放时间越长. 实现 PSC 映射的步骤如下:

**步骤 1** 将质心的水平方向坐标  $x_c$  和垂直方向坐标  $y_c$  均匀量化为三个等级, 分别赋值为 0, 1, 2.

**步骤 2** 统计 ROI 区域内像素点的个数  $S$ , 用来控制音符序列的长度.

**步骤 3** 根据公式 (19) 生成数字音符序列.

$$y(i) = 2^{x_c} \sin\left(\frac{i \times 2\pi \times 2^{y_c} \times f_r}{F_s}\right), i = 1, 2, \dots, S \times \delta \quad (19)$$

其中  $x_c, y_c$  表示量化后的质心坐标;  $f_r$  表示由不同颜色映射的音符的频率;  $2^{x_c}$  为响度调节因子;  $2^{y_c}$  为音高调节因子;  $F_s$  为采样频率;  $\delta$  为音符序列的最小长度.

## 4 实验结果与分析

为了验证基于注意模型的视觉替代方法的有效

性, 在 Intel Core(TM)2.80GHz 处理器的 PC 机环境下, 采用室内外场景图像进行实验. 实验分 4 步进行.

**实验 1** 提取实验图像的 ROI 区域, 并计算位置、尺寸和主颜色特征. 首先, 使用华盛顿大学的自然景物图像库 (Ground Truth Database) 测试文中提出的 ROI 区域主颜色提取算法—FDH 算法的正确率. 当算法提取的主颜色与人眼判断的主颜色一致时, 认为算法结果是正确的, 否则是错误的. 实验结果如表 2 所示. 结果表明: FDH 算法在提取 ROI 区域主颜色时的正确率明显高于直接应用颜色直方图法提取主颜色的正确率. 而增加膨胀运算的次数, 还可以提高 FDH 算法的正确率, 但是会增加运算时间. 当膨胀次数为 3 次时, 正确率达到 96% 以上, 已经能够满足系统的需求.

表 2 FDH 算法的正确率

图像集	图像数	ROI 数	颜色直方图法	FDH 算法
Japan	20	96	89.6%	97.9%
Green land	20	88	87.5%	96.6%
Yellow stone	20	75	89.3%	98.7%

然后用大小为  $256 \times 384$  的室内场景图像进行实验, 实验结果如图 2 所示, 图 2(a) 为彩色输入图像, 图 2(b) 中的白色区域为提取的三个 ROI 区域, 图 2(c) 是 166 色颜色量化后的 ROI 区域的图像, 图中出现了量化误差造成的颜色突变点; 图 2(d) 是颜色滤波后的图像, 去除了颜色突变点和像素数目较少的颜色. 由于 ROI 区域 3 中的目标物比较小, 因此直接使用颜色直方图法提取的主颜色是背景的颜色, 经过 3 次膨胀后, 目标物的颜色增加, 此时再使用颜色直方图提取主颜色, 可以获得正确的结果. ROI 区域的主颜色、位置、尺寸特征如表 3 所示.

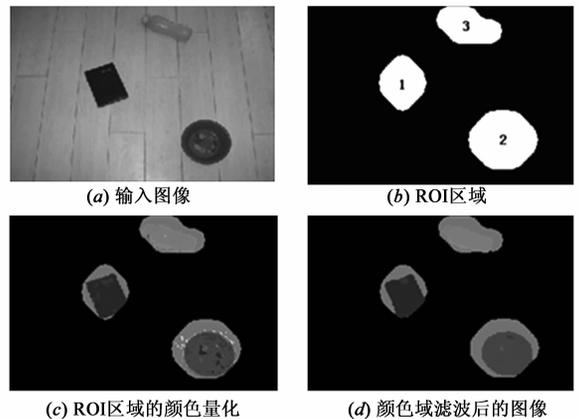


图2 图像的ROI区域

**实验 2** 将 ROI 区域的位置、尺寸和颜色特征映射为音符. 映射结果如表 4 所示. 三个 ROI 区域分别位于不同的垂直位置上, 因此映射为三种不同的音高. 而 ROI 区域 1 和 3 的水平位置都处于图像的中部, 所以两

者映射的响度相同. 音符的播放时间是在 44.1KHz 的采样频率下测得. ROI 区域 1 和 3 的面积近似相等, 因此两者映射的音长近似相等. 经过 PSC 映射后, 响度从左向右增高; 而音高从上至下升高. PSC 映射方法与文献 1 中提出的映射方法相比, 去除了图像中的次要信息后映射为电子音符, 降低了盲人学习训练的难度, 且听感良好, 长时间使用不会造成听觉疲劳; 与文献 2 中提出的映射方法相比, 增加了颜色和尺寸特征的映射, 能帮助盲人获得更多的外界环境信息.

表 3 ROI 区域的颜色、位置、尺寸特征

ROI 区域	主颜色	质心	尺寸
1	黑色	(135, 108)	3704
2	红色	(278, 189)	6364
3	绿色	(231, 33)	3770

表 4 ROI 区域的映射结果

ROI 区域	音符	响度(振幅)	音长/秒
1	B5	2	0.4200
2	# C6	4	0.7215
3	G4	2	0.4274

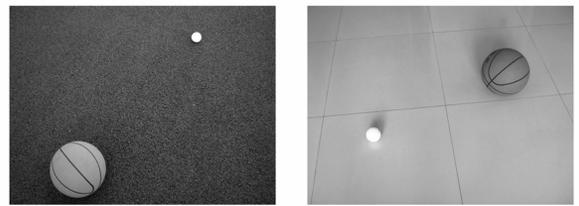
**实验 3** 对实验对象进行听音训练与测试, 验证算法的效果. 实验对象 1, 22 岁, 非盲人, 听力正常, 实验时用眼罩遮住双眼. 实验对象 2, 46 岁, 后天盲人, 失明 24 年, 听力正常. 首先, 用映射算法可能产生的电子音符进行听音训练, 并学习 ROI 区域特征与音符之间的映射规则. 然后, 用 20 幅室内外场景图像(包含 ROI 区域 72 个)进行实验, 记录实验对象通过听音确定的 ROI 区域的位置、尺寸和颜色, 并与实际的位置、尺寸和颜色相比较, 确定实验对象听音测试的准确率, 结果如表 5 所示.

表 5 实验对象听音测试的准确率

实验对象	颜色	位置	面积
1	88.9%	93.1%	83.3%
2	94.4%	97.2%	86.1%

结果表明: 实验对象 2(盲人)通过听音判断信息的准确率要高于实验对象 1(非盲人). 而两者对于 ROI 区域颜色和位置判断的准确率要高于对面积判断的准确率. 这是因为颜色映射为十二平均率设定的标准音符音调, 人耳很容易分辨; 而位置坐标经量化后映射为成指数增加的音高和响度, 人耳也比较容易分辨. 面积直接映射为音长, 这是为了避免量化误差可能将相似尺寸的 ROI 区域映射为完全不同的音长. 通过增加学习训练的次数, 人耳对音长的分辨率会进一步提高. 经过简单的学习训练之后, 盲人能很好的应用该方法获得图像 ROI 区域的主要信息.

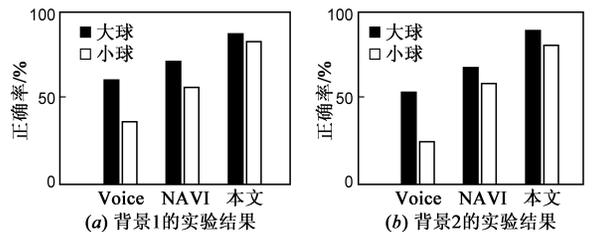
**实验 4** 通过在不同背景下搜索大球和小球的“Searching Balls”实验, 验证算法的优越性. 分别以柏油路面(背景 1)、乳白色瓷砖地面(背景 2)为背景, 在随机位置上放置一个大球(篮球)和一个小球(网球), 如图 3 所示, 并不断改变两球所在位置, 拍摄 100 张实验图像. 其中 50 张图像构成训练集, 另外 50 张图像构成测试集. 分别采用 Voice 系统、NAVI 系统和本文提出的方法将训练集中的图像映射为声音, 盲人收听声音, 并由工作人员向盲人解释该图像中大球和小球所在的位置, 循环进行 10 次训练, 帮助盲人在大脑中建立一种声音和大球、小球位置的对应关系. 然后用测试集进行实验, 盲人通过收听三种方法产生的声音, 指出大球和小球在图像中的位置. 实验结果如图 4 所示.



(a) 背景1

(b) 背景2

图3 “Searching Balls”实验



(a) 背景1的实验结果

(b) 背景2的实验结果

图4 “Searching Balls”实验结果

结果表明: 在背景 1 和背景 2 中, 本文方法搜索大球和小球位置的正确率明显高于 Voice 和 NAVI 系统. Voice 和 NAVI 系统以灰度图像为映射对象. 将背景 1 图像变为灰度图像后, 大球和小球灰度值较高, 而柏油路面灰度值较低. 将背景 2 图像变为灰度图像后, 小球和瓷砖地面具有较高灰度值, 大球具有较低灰度值. 根据 Voice 系统映射算法, 灰度值直接映射为声音信号的振幅. 背景 2 中的瓷砖地面映射为声音后形成较强的干扰, 因此, Voice 算法在背景 2 中的正确率低于背景 1. NAVI 系统首先识别出图像中的目标物, 然后提高目标物的灰度级、降低背景的灰度级, 从而减少背景信息的干扰. 而本文提出的方法, 利用了视觉注意模型来选择映射区域, 受背景影响较小, 并且 PSC 映射方法将映射区域的特征映射为更容易学习和使用的电子音符, 降低了盲人使用的难度, 因此在“Searching Balls”实验中取得了较高的正确率.

## 5 结论

本文将视觉注意模型应用到盲人视觉替代问题中,解决了视觉替代研究中无法为各种不同的应用场景建立统一的目标识别模型的问题.针对人眼的视觉认知特性,提出了将 ROI 区域的位置、面积和颜色映射为音符响度、音长和音调的 PSC 映射方法.这种视觉替代方法可以帮助盲人在各种不同的环境中获得重要的外界信息.通过对这些信息的分析和理解,盲人可以安全有效地完成行走、寻物等任务.

### 参考文献

- [1] Meijer P B L. An experimental system for auditory image representations[J]. IEEE Transactions on Biomedical Engineering, 1992, 39 (2): 112 - 121.
- [2] Sainarayanan G, Nagarajan R, Yaacob S. Fuzzy image processing scheme for autonomous navigation of human blind[J]. Applied Soft Computing, 2007, 7(1): 257 - 264.
- [3] Bologna G, Deville B, Pun T. On the use of the auditory pathway to represent image scenes in real-time[J]. Neurocomputing, 2009, 72(4): 839 - 849.
- [4] Chen C L, Liao Y F, Tai C L. Image to MIDI mapping based on dynamic fuzzy color segmentation for visually impaired people[J]. Pattern Recognition Letters, 2011, 32(4): 549 - 560.
- [5] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20 (11): 1254 - 1259.
- [6] Siagian C, Itti L. Rapid biologically-inspired scene classification using features shared with visual attention[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29 (2): 300 - 312.

- [7] 王正友, 胡国胜, 明建华, 等. 基于 SSIM 与 ROI 的视频质量评价方法[J]. 仪器仪表学报, 2009, 30(9): 1906 - 1911.  
Wang Zhengyou, Hu guosheng, Ming Jianhua, et al. Video quality assessment based on SSIM and ROI[J]. Chinese Journal of Scientific Instrument, 2009, 30(9): 1906 - 1911. (in Chinese)
- [8] 黄凯奇, 王桥, 吴镇杨. 基于视觉特性和颜色空间的多尺度彩色图像增强算法[J]. 电子学报, 2004, 32(4): 673 - 676.  
Huang Kaiqi, Wang Qiao, Wu Zhenyang. Multi-scale color image enhancement algorithm based on color space and human visual system (HVS)[J]. Acta Electronica Sinica, 2004, 32(4): 673 - 676. (in Chinese)
- [9] Smith J R. Integrated Spatial and Feature Image Systems: Retrieval, Analysis and Compression [D]. Columbia University, 1997.
- [10] Berlin B, Kay P. Basic Colors Terms: Their Universality and Evolution[M]. Berkeley, CA: University of California Press, 1991. 15 - 25.

### 作者简介



**田亚男** 女, 1979 年生于辽宁阜新. 现为东北大学信息科学与工程学院讲师, 博士研究生. 研究方向为智能信息处理, 图像处理, 模式识别等. E-mail: tianyanan@ise.neu.edu.cn

**雷红玮** 女, 1975 年生于河南. 现为东北大学信息科学与工程学院讲师. 研究方向为生物信息检测.